第四章：生成排列和组合

4.1 生成排列
4.2 排列中的逆序
4.3 生成组合
4.4 生成 r 子集

4.4 生成 *r*子集算法

例. 生成{4, 3, 2, 1}的所有2子集

方法一：

效率低！

1. 生成所有组合

0000	0001	0010	$\boxed{0011}$	0100	$\boxed{0101}$	$\boxed{0110}$	0111
1000	$\boxed{1001}$	$\boxed{1010}$	1011	$\boxed{1100}$	1101	1110	1111

2. 选出所有2子集

能否直接生成所有2子集？

{2, 1}, {3, 1}, {3, 2}, {4, 1}, {4, 2}, { 4, 3} *r*子集的字典序：

- 令 $S=\{1, 2, \ldots, n\}$由前 n 个正整数组成。
  - 给出 S 的元素的一个自然顺序：

$$1 < 2 < \ldots < n$$

属于A或B，但不同时属于A和B

- 设A，B是 S 的两个 r 组合，若 $A\cup B\backslash A\cap B$ 中的最小整数属于A，则称 A先于B。

例 $\{1, 2, 3, 4, 5, 6, 7, 8\}$的两个5子集

$A=\{2，3，4，7，8\}$，$B=\{2，3，5，6，7\}$

$A\cup B\backslash A\cap B=\{4, 5, 6, 8\}$

A以字典序先于B

- 组合表示为子序列
  - 约定 $S=\{1, 2, \ldots, n\}$ 的 r 子集为如下形式：

$$a_1 a_2 \ldots a_r,\ \text{其中}\ 1 \leq a_1 < a_2 < \ldots < a_r \leq n$$

例： $\{1, 2, 3, 4, 5, 6, 7, 8\}$的两个 5子集：

23478 先于 23567

若 $A \cup B \backslash A \cap B$ 中的最小整数属于A，则称 A先于B。

- 组合表示为子序列
  - 约定 $S=\{1, 2, \ldots, n\}$ 的 r 子集为如下形式：

$a_1 a_2 \ldots a_r$，其中 $1 \leq a_1 < a_2 < \ldots < a_r \leq n$

例: $S=\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$的 5子集按字典序排序，则

第一个是: 12345 ，最后一个: 56789

12589的直接后继是 12678

- 以 1258 开头的后继：因为 12589 是最后一个，所以无后继
- 以 125 开头的后继：因为 12589 是最后一个，所以无后继
- 以 12 开头的后继：12678 为第一个
- 以 1 开头的后继：13456 为第一个
- 第 1 位比 1 大的后继：23456 为第一个

12467 的直接后继是12468， 24679 的直接后继是 24689

- 设 $S=\{1, 2,\ldots, n\}$, $a_1\ldots a_r$是 S 的一个 r子集。

  (1) 对任意的$1\leq i<j\leq r-1$,

  $a_1\ldots a_r$ 的以 $a_1\ldots a_j$开头的第一个后继一定先于以 $a_1\ldots a_i$开头的第一个后继（如果存在）。

例: $S=\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$的 5子集按字典序排序，考虑 12467 的直接后续。

- 以 1246 开头的后继: 12468 是第一个
- 以 124 开头的后继： 12478 是第一个
- 以 12 开头的后继： 12567 是第一个
- 以 1 开头的后继： 13456 是第一个
- 第 1 位比 1 大的后继：23456 是第一个

- 设 $S=\{1, 2,\ldots, n\}$, $a_1\ldots a_r$是 S 的一个 r子集。

  (1) 对任意的$1\leq i<j \leq r-1$,

  $a_1\ldots a_r$ 的以 $a_1\ldots a_j$ 开头的第一个后继一定先于以 $a_1\ldots a_i$ 开头的第一个后继（如果存在）。

记以 $a_1\ldots a_j$ 开头的第一个后继:

$$a_1\ a_2\ldots a_i\ \ a_{i+1}\ \ldots\ a_j\ b_{j+1}\ldots\ b_r$$

以 $a_1\ldots a_i$ 开头的第一个后继:

$$a_1\ a_2\ \ldots\ a_i\ c_{i+1}\ \ldots\ c_r$$

一定有: $a_{i+1}<c_{i+1}<\ldots<c_r$, 且 $a_{i+1}< a_{i+2}<\ldots\ a_j < b_{j+1}<\ldots< b_r$

令$A=\{a_1, a_2, \ldots, a_i,\ a_{i+1}, \ldots, a_j, b_{j+1}, \ldots, b_r\}$,

$B=\{a_1, a_2, \ldots , a_i, c_{i+1}, \ldots , c_r\}$,

则 a_{i+1}一定是$A\cup B\backslash A\cap B$中的最小数，因此(1)成立。

- 设 $S=\{1, 2,\ldots, n\}$, $a_1\ldots a_r$是 S 的一个 r子集。

(1) 对任意的$1\leq i < j \leq r\text{-}1$，

$a_1\ldots a_r$ 的以 $a_1\ldots a_j$ 开头的第一个后继一定先于以 $a_1\ldots a_i$ 开头的第一个后继（如果存在）。

(2) 如果$a_r< n$, 则 $a_1\ldots a_r$ 的直接后继为$a_1\ldots a_{r-1}\ a_r+1$.

以 $a_1\ldots a_{r-1}$ 开头的第一个后继

定理4.4.1 (1) 设 $a_1a_2\ldots a_r$ 是 $\{1, 2, \ldots, n\}$ 的 r 子集。在字典序中，第一个 r子集是 $12\ldots r$，最后一个 r子集是 $(n-r+1)(n-r+2)\ldots n$。

$a_1\ldots a_{r-1}$ 不是最后一个r子集

(2) 设$a_1a_2\ldots a_r \neq (n-r+1)(n-r+2)\ldots n$。令 k 是满足 $a_k < n$且使得a_k+1不同于$a_1a_2\ldots a_r$中任一数的最大整数。那么, 在字典序中, $a_1a_2\ldots a_r$ 的直接后继是

$$a_1a_2\ldots a_{k-1}(a_k+1)(a_k+2)\ldots(a_k+r-k+1).$$

例如：$n=9$, $r=6$, $k=4$, $a_1a_2\ldots a_6$ 的直接后继是 $a_1a_2a_3(a_4+1)(a_4+2)(a_4+3)$.

定理4.4.1 (1) 设 $a_1a_2\ldots a_r$ 是 $\{1, 2, \ldots, n\}$ 的 r 子集。
在字典序中，第一个 r子集是 $12\ldots r$，最后一个 r子集是 $(n-r+1)(n-r+2)\ldots n$。

$a_1\ldots a_{r-1}$ 不是最后一个r子集

(2) 设$a_1a_2\ldots a_r \neq (n-r+1)(n-r+2)\ldots n$。令 k 是满足 $a_k < n$且使得a_k+1不同于$a_1a_2\ldots a_r$中任一数的最大整数。那么, 在字典序中, $a_1a_2\ldots a_r$ 的直接后继是

$$a_1a_2\ldots a_{k-1}(a_k+1)(a_k+2)\ldots(a_k+r-k+1).$$

直接后继求解算法:

1. 当$a_i<n$时$(1\leq i\leq r)$, 求a_i+1, 判断a_i+1是否属于$\{a_1, \ldots, a_r\}$;
2. 找出满足$a_i < n$, 且a_i+1不在$\{a_1,\ldots, a_r\}$中的最大的 i, 记为 k,

那么, 在字典序中, $a_1a_2\ldots a_r$ 的直接后继是

$$a_1a_2\ldots a_{k-1}(a_k+1)(a_k+2)\ldots(a_k+r-k+1)$$

1. 当$a_i<n$时$(1\le i \le r)$, 求a_i+1, 判断a_i+1是否属于$\{a_1, \ldots, a_r\}$;

2. 找出满足$a_i<n$, 且a_i+1不在$\{a_1,\ldots, a_r\}$中的最大的 i, 记为 k,

那么, 在字典序中, $a_1a_2\ldots a_r$ 的直接后继是

$$a_1a_2\ldots a_{k-1}\ (a_k+1)\ (a_k+2)\ldots(a_k+r-k+1)$$

例. 考虑{1, 2, 3, 4, 5, 6, 7, 8}的5子集:

$a_1a_2a_3a_4a_5$

5子集 1 3 4 5 7的直接后继是 13458

a_i+1: 2 4 5 6 8, k=5

5子集 13578的直接后继是 13678

a_i+1: 2468, k=3

5子集 12458的直接后继是 12467

a_i+1: 2356, k=4

定理4.4.1 (1) 令$a_1a_2\ldots a_r$是$\{1, 2, \ldots, n\}$的一个 r 组合, 在字典序中, 第一个r子集是$12\ldots r$, 最后一个r 组合是 $(n-r+1)(n-r+2)\ldots n$。

(2) 设$a_1a_2\ldots a_r \neq (n-r+1)(n-r+2)\ldots n$。令 k 是满足 $a_k<n$且使得a_k+1不同于$a_1a_2\ldots a_r$中任一数的最大整数。

那么, 在字典序中, $a_1a_2\ldots a_r$的直接后继是

$$a_1a_2\ldots a_{k-1}(a_k+1)(a_k+2)\ldots(a_k+r-k+1).$$

证明: (1) 根据字典序的定义知，第一个组合是$12\ldots r$, 最后一个是的$(n-r+1)(n-r+2)\ldots n$。

(2) 设 $a_1a_2\ldots a_r$不是最后一个 r子集，k 为使得$a_k<n$且a_k+1不同于$a_1a_2\ldots a_r$中任一数的最大整数。

下面分两种情况进行证明： $a_r<n$ 或 $a_r=n$

定理4.4.1 (1) 令$a_1a_2\ldots a_r$是$\{1, 2, \ldots, n\}$的一个 r 组合, 在字典序中, 第一个r子集是$12\ldots r$, 最后一个r 组合是 $(n-r+1)(n-r+2)\ldots n$。

(2) 设$a_1a_2\ldots a_r \neq (n-r+1)(n-r+2)\ldots n$。令 k 是满足 $a_k<n$且使得a_k+1不同于$a_1a_2\ldots a_r$ 中任一数的最大整数。

那么, 在字典序中, $a_1a_2\ldots a_r$ 的直接后继是

$$a_1a_2\ldots a_{k-1}(a_k+1)(a_k+2)\ldots(a_k+r-k+1).$$

证明： (a) $a_r<n$时， 显然 $a_1a_2\ldots a_r$ 的字典序直接后继是以$a_1a_2\ldots a_{r-1}$开头的第一个后继，即 $a_1\ldots a_{r-1}(a_r+1)$。

因为$a_r<n$时, a_r+1肯定不同于$a_1a_2\ldots a_r$中任一数，且r是满足条件的最大整数,

则由定理4.4.1生成的后继为$a_1\ldots a_{r-1}(a_r+1)$，与前面一致。

定理4.4.1(1) $a_1a_2\ldots a_r$是$\{1,2,\ldots,n\}$的一个r子集, 在字典序中, 第一个r子集是$12\ldots r$, 最后一个r子集是$(n-r+1)\ (n-r+2)\ldots n$。

(2)设 $a_1a_2\ldots a_r \neq (n-r+1)\ (n-r+2)\ldots n$。令 k 是满足 $a_k<n$ 且使得a_k+1不同于$a_1a_2\ldots a_r$中任一数的最大整数。那么, 在字典序中, $a_1a_2\ldots a_r$ 的直接后继是 $a_1a_2\ldots a_{k-1}\ (a_k+1)\ldots(a_k+r-k+1)$.

证明：(b) 当$a_r=n$时，假设有满足条件的k，此时有 $k\neq r$。

共r-k-1项

a_1	a_2	…	a_k	a_{k+1}	a_{k+2}	a_{k+3}	…	a_{r-1}	$a_r=n$
a_1+1	a_2+1	…	a_k+1	$a_{k+1}+1$	$a_{k+2}+1$	$a_{k+3}+1$	…	$a_{r-1}+1$	$a_r=n$

共r-k-1项 $>a_k+1$

则必有 $a_{k+2}=a_{k+1}+1,\ a_{k+3}=a_{k+2}+1,\ \ldots,\ a_{r-1}=a_{r-2}+1,\ a_r=a_{r-1}+1=n$。

得$a_1\ldots a_r = a_1\ldots a_k,\ a_{k+1},\ a_{k+1}+1,\ a_{k+1}+2,\ \ldots,\ a_{k+1}+(r-k-1) = a_r = n$

因此， $a_{k+1}=n-(r-k-1)$， $a_k+1< a_{k+1}$。

从而， $a_1\ldots a_r = a_1\ldots a_k,\ n-(r-k-1),\ n-(r-k-2),\ldots,\ n$.

定理4.4.1(1) $a_1a_2\ldots a_r$是$\{1,2,\ldots,n\}$的一个r子集, 在字典序中, 第一个r子集是$12\ldots r$, 最后一个r子集是$(n-r+1)\ (n-r+2)\ldots n$。(2)设$a_1a_2\ldots a_r \neq (n-r+1)\ (n-r+2)\ldots n$。令$k$是满足$a_k<n$且使得$a_k+1$不同于$a_1a_2\ldots a_r$中任一数的最大整数。那么, 在字典序中, $a_1a_2\ldots a_r$ 的直接后继是 $a_1a_2\ldots a_{k-1}\ (a_k+1)\ldots(a_k+r-k+1)$.

证明：(b) 当$a_r=n$时，假设有满足条件的k，此时有 $k\neq r$。

得到 $a_1\ldots a_r = a_1\ldots a_k, a_{k+1}, a_{k+1}+1, a_{k+1}+2, \ldots, a_{k+1}+(r-k-1)$

$= a_1\ldots a_k, n-(r-k-1), n-(r-k-2), \ldots, n$.

其中，$a_k+1< a_{k+1}=n-r+k+1$。

因此，$a_1a_2\ldots a_r$ 是以$a_1\ldots a_{k-1}a_k$ 开始的最后的 r 子集。

而$a_1\ldots a_{k-1}(a_k+1)(a_k+2)\ldots(a_k+r-k+1)$是以

$a_1\ldots a_{k-1}(a_k+1)$开始的第一个 r 子集，结论成立。

$\{1, 2, \ldots, n\}$的字典序 r 子集的生成算法

- 从$12\ldots r$ 开始，逐个列出直接后继，直至得到$(n-r+1)(n-r+2)\ldots n$

1. 初始： $a_1a_2\ldots a_r = 12\ldots r$；

2. 当$a_1a_2\ldots a_r \neq (n-r+1)(n-r+2)\ldots n$ 时，进行以下操作：

   (1) 确定最大整数k, 使得

   $$a_k+1 \leq n，且 a_k+1 \neq a_i \ (\mathrm{i}=1,2,\ldots,\mathrm{r});$$

   (2) 用$a_1a_2\ldots a_{k-1}(a_k+1)\ldots(a_k+r-k+1)$替换 $a_1a_2\ldots a_r$.

例：应用算法生成{1,2,…,6}的所有4子集

a_i+1： 2345 2346 234 2356 235

1234 → 1235 → 1236 → 1245 → 1246

a_i+1： 236 2456

→1256 →1345 →1346→1356→1456

→2345→2346→2356→2456→ 3456

例： 生成{1, 2, 3, 4, 5, 6, 7}的所有5-元组

定理4.2.2 $\{1, 2, \ldots, n\}$的r子集 $a_1a_2\ldots a_r$ 出现在 $\{1, 2, \ldots, n\}$ 的 r 子集中的字典序中的位置号为：

$$\binom{n}{r}-\binom{n-a_1}{r}-\binom{n-a_2}{r-1}-\cdots-\binom{n-a_{r-1}}{2}-\binom{n-a_r}{1}$$

证明：位置号为所有r组合的个数减去其所有后继的个数。首先计算$a_1a_2\ldots a_r$的所有后继的个数。分两种情况：

(1) 当$1 \le i \le r-1$时，设以$a_1a_2\ldots a_i$开头的 $a_1a_2\ldots a_r$的后继为 $a_1a_2\ldots a_i b_1\ldots b_{r-i}$，则有：

$$a_{i+1} < b_1 < \ldots < b_{r-i} \le n。$$

因此，以$a_1a_2\ldots a_i$开头的$a_1a_2\ldots a_r$的后继个数为：

$\binom{n-a_{i+1}}{r-i}$, $i=1,\ldots, r-1$

(2) 第一个位置比a_1大的后继的个数为$\binom{n-a_1}{r}$

因此，可得位置号。证毕。

定理4.2.2 {1, 2,…, n}的r子集 $a_1a_2\ldots a_r$ 出现在 {1, 2, …, n} 的 r 子集中的字典序中的位置号为:

$$\binom{n}{r}-\binom{n-a_1}{r}-\binom{n-a_2}{r-1}-\cdots-\binom{n-a_{r-1}}{2}-\binom{n-a_r}{1}$$

例. 求{ 1, 2, …, 8 }的 4子集 1258 的字典序位置。

解：1258的位置是:

$$\binom{8}{4}-\binom{7}{4}-\binom{6}{3}-\binom{3}{2}-\binom{0}{1}=22$$

第四章 内容小节

- 排列生成算法：
  - 递归生成算法
  - 邻位对换算法
  - 从逆序生成排列
- 组合生成算法
  - 字典序
  - 反射**Gray**码
  - 基于字典序的 r-子集生成算法

《计算机程序设计艺术》第**4**卷**2**册——生成所有元组和排列，**Donald E. Knuth**著，苏运霖译。

Donald E. Knuth　高德纳

- **1938年1月10日生于美国威斯康星州密尔沃基市**
- 他的超凡智力在8岁时就显示出来了，用“**Ziegler’s Giant Bar**”里面的字母，写单词。裁判准备了一份**2500**个单词的列表，而高德纳却写出了**4500**多个单 词，获得了冠军。他的赛后感言是“我还能写出更多”。
- **1960年Donald E. Knuth 22**岁毕业，由于“成绩过于优异”，同时被授予学士和硕士学位。
- 他在**36**岁的时候就获得了图灵奖（**Unix**的发明人之一**Ken Thompson** 是到**40**多岁才拿图灵奖的)
- **Knuth**总共教了**28**个博士生。不知道怎么搞的，他觉得**28**这个数字很好，于是就决定再也 不带博士生了。

- **1962年，世界上一流的出版社Addison-Wesley艾迪生-韦斯利出版社约初露头角的高德纳写一本编译器和程序设计方面的书，这件原本寻常的事最终成就了计算机科学史上的一个奇观。**
- **1962年约的稿，高德纳一直写到1966年还没交(写了4年），编辑找到高德纳，说这都四年了你写了多少啊，高德纳说，才写3000页手稿。编辑大囧，忙问都3000页了你怎不交，高德纳答曰，急啥，我还没写到正题呢。编辑彻底雷住了，说那你出个多卷本吧……**

《计算机程序设计艺术》，就这么诞生了，计划写7卷。

- 1968年，《计算机程序设计艺术》(TAOCP)的第一卷：基本算法正式出版了。
- 微软首席执行官比尔盖茨在1995年接受一次采访时说，“如果你认为你是一名真正优秀的程序员，就去读第一卷，确定可以解决其中所有的问题。”盖茨本人读这本书时用去了几个月的时间，并同时进行了难以置信的训练。
- 盖茨还说：“如果你能读懂整套书的话，请给我发一份你的简历。”高德纳本人的说法更犀利：要是看不懂，就别当 程序员。
- 1970年第二卷半数值算法出版，1973年第三卷排序与查找出版，这三卷书立即被计算机界惊为神作，在那几年就卖出去100多万套，至今 仍是编程书籍中的最高经典。

- 1974年，高纳德获得图灵奖，保持着获奖年龄最小的纪录

BIRTH:

January 10, 1938, in Milwaukee, Wisconsin.

EDUCATION:

Graduated from Milwaukee Lutheran High School (1956); BS in mathematics from the

DONALD ("DON") ERVIN KNUTH

United States – 1974

CITATION

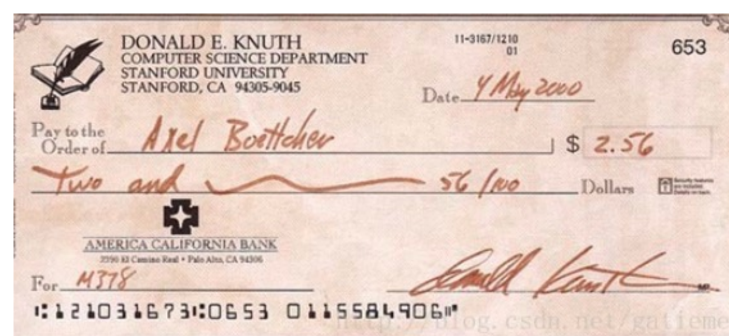
For his major contributions to the analysis of algorithms and the design of programming languages, and in particular for his contributions to the "art of computer programming" through his well-known books in a continuous series by this title.

- 封笔十年，创造了三个重要的成果：
  - 排版系统TEX
  - 字体设计系统METAFONT
  - 文学化编程(Literate Programming)

谁发现TEX的一个错误，就付他2.56 美元，第二个错误5.12美元，第三个10.24美元……以此类推。

另 一个奖项是找出其著作中错误的人能得到2.56美元，因为“256美分刚好是十六进制的一美元”

有网友戏说，什么是聪明：在 Knuth 的书中找到错误；什么是愚蠢：去兑现那张两块五毛六的支票。



- **2008年，在TAOCP的前三卷面市30年之后，第四卷终于面世了**

*正如当年**Linux**的作者**Linus**说：上帝在梦中告诉我，我做出了最优秀的操作系统。高德纳回答说：我可没这么说过*